**GROUP ID: GE04**

**A PROJECT REPORT ON**

# Automated Multi-Stage Blood Cancer Detection Using Deep Learning

SUBMITTED TO THE PIMPRI CHINCHWAD COLLEGE OF ENGINEERING AN AUTONOMOUS INSTITUTE, PUNE
IN THE FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

**BACHELOR OF TECHNOLOGY**
**COMPUTER ENGINEERING (REGIONAL LANGUAGE)**

**SUBMITTED BY**


| | |
|---|---|
| SEJAL KOTKAR | PRN: 121B1D033 |
| CHARUL CHIM | PRN: 121B1D047 |
| SANIKA MENKUDALE | PRN: 122B2D077 |




**DEPARTMENT OF COMPUTER ENGINEERING (REGIONAL LANGUAGE)**

**PCET'S PIMPRI CHINCHWAD COLLEGE OF ENGINEERING**

**Sector No. 26, Pradhikaran, Nigdi, Pimpri-Chinchwad, PUNE 411044**

**2024-2025**

# CERTIFICATE

This Is To Certify That The Project Report Entitles

**"Automated Multi-Stage Blood Cancer Detection Using Deep Learning"**

Submitted by

| | |
|---|---|
| SEJAL KOTKAR | PRN: 121B1D031 |
| CHARUL CHIM | PRN: 121B1D055 |
| SANIKA MENKUDALE | PRN: 122B2D077 |

is a Bonafide student at this institute and the work has been carried out by her under the supervision of **Prof. ANANDKUMAR BIRAJDAR,** and it is approved for the partial fulfillment of the requirement of Pimpri Chinchwad College of Engineering an autonomous institute, for the award of the B. Tech. degree in Computer Engineering.

**(Prof. Anandkumar Birajdar)**
Project Guide

**(Prof. Dr. Rachana Y. Patil)**
Head of Department

**(Prof. Dr. G.N. Kulkarni)** Director,
Pimpri Chinchwad College of Engineering Pune – 44

**Place:** Pune
**Date:**

# CERTIFICATE

This Is To Certify That The Project Report Entitles

**“Automated Multi-Stage Blood Cancer Detection Using Deep Learning”**

Submitted by

Student Name: **Sejal Kotkar** PRN: **121B1D031**

is a bonafide student of this institute and the work has been carried out by her under the supervision of **Prof. ANANDKUMAR BIRAJDAR** and it is approved for the partial fulfillment of the requirement of Pimpri Chinchwad College of Engineering an autonomous institute, for the award of the B. Tech. degree in Computer Engineering.

**(Prof. Anandkumar Birajdar)**
Project Guide

**(Prof. Dr. Rachana Y. Patil)**
Head of Department

**(Prof. Dr. G.N. Kulkarni)** Director,
Pimpri Chinchwad College of Engineering Pune – 44

**Place:** Pune
**Date:**

# CERTIFICATE

This Is To Certify That The Project Report Entitles

**“Automated Multi-Stage Blood Cancer Detection Using Deep Learning”**

Submitted by

Student Name: **Charul Chim** PRN: **121B1D055**

is a bonafide student of this institute and the work has been carried out by him under the supervision of **Prof. ANANDKUMAR BIRAJDAR** and it is approved for the partial fulfillment of the requirement of Pimpri Chinchwad College of Engineering an autonomous institute, for the award of the B. Tech. degree in Computer Engineering.

**(Prof. Anandkumar Birajdar)**
Project Guide

**(Prof. Dr. Rachana Y. Patil)**
Head of Department

**(Prof. Dr. G.N. Kulkarni)** Director,
Pimpri Chinchwad College of Engineering Pune – 44

**Place:** Pune
**Date:**

# CERTIFICATE

This Is To Certify That The Project Report Entitles

**“Automated Multi-Stage Blood Cancer Detection Using Deep Learning”**

Submitted by

Student Name: **Sanika Menkudale** PRN: **121B1D077**

is a bonafide student of this institute and the work has been carried out by him under the supervision of **Prof. ANANDKUMAR BIRAJDAR** and it is approved for the partial fulfillment of the requirement of Pimpri Chinchwad College of Engineering an autonomous institute, for the award of the B. Tech. degree in Computer Engineering.

**(Prof. Anandkumar Birajdar)**
Project Guide

**(Prof. Dr. Rachana Y. Patil)**
Head of Department

**(Prof. Dr. G.N. Kulkarni)** Director,
Pimpri Chinchwad College of Engineering Pune – 44

**Place:** Pune
**Date:**

# ACKNOWLEDGEMENT

We express our sincere thanks to our **Guide Prof. Anandkumar Birajdar** for his constant encouragement and support throughout our project, especially for the useful suggestions given during the course of the project and having laid down the foundation for the success of this work.

We would also like to thank our **Project Coordinator, Prof. Rohini Y. Sarode** for her assistance, genuine support and guidance from early stages of the project. We would like to thank **Prof. Dr. Rachana Y. Patil, Head of Computer Engineering (Regional Language) Department** for her unwavering support during the entire course of this project work. We are very grateful to our **Director, Prof. Dr. G.N. Kulkarni** for providing us with an environment to complete our project successfully. We also thank all the staff members of our college and technicians for their help in making this project a success.

We also thank all the web committees for enriching us with their immense knowledge. Finally, we take this opportunity to extend our deep appreciation to our family and friends, for all that they meant to us during the crucial times of the completion of our project.

| NAME OF THE STUDENTS | SIGN |
|---|---|
| SEJAL KOTKAR | |
| CHARUL CHIM | |
| SANIKA MENKUDALE | |

# ABSTRACT

Blood cancer remains a critical global health issue due to its high mortality rate, complex treatment procedures, and challenges in early and accurate diagnosis. Traditional diagnostic methods rely on manual examination of microscopic blood smear images, which is time-consuming, labour-intensive, highly subjective, and prone to human error. With advancements in deep learning, artificial intelligence (AI), and medical imaging, automated systems can significantly enhance diagnostic efficiency, accuracy, and early detection, leading to improved patient outcomes and timely medical intervention.

In this Project, we propose an AI-driven deep learning model based on MobileNet for automated blood cancer detection and precise stage classification. Our approach involves preprocessing microscopic blood smear images, including noise removal, contrast enhancement, morphological processing, edge detection, and feature extraction, to improve detection precision and ensure high-quality input data. The dataset used consists of thousands of labelled blood cell images, categorized into four progressive cancer stages: Benign, Malignant Early, Malignant Pre, and Malignant Pro, allowing for an efficient and reliable classification system.

We use transfer learning with a pre-trained MobileNet model, adding fully connected layers for cancer stage classification. The dataset is split into training (70%), validation (15%), and testing (15%) sets, with data augmentation (zooming, shearing, rotation, flipping) to enhance generalization. Trained using categorical cross-entropy and RMSprop, the model achieves 92–95% accuracy. MobileNet proves to be an efficient, lightweight solution for real-world medical use. Future work includes model optimization, expanding dataset diversity, and integrating it into a web-based diagnostic system.

**Keywords:** Blood Cancer Detection, Deep Learning, MobileNet, Transfer Learning, Microscopic Blood Smear Images, Feature Extractions, AI in Healthcare, Medical Image Processing, Cancer Stage Classification, Data Augmentation

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| ABBREVIATION | ILLUSTRATION |
|---|---|
| AUC | Area Under the Curve |
| CV | Cross-Validation |
| ROC | Receiver Operating Characteristic |
| XGBoost | eXtreme Gradient Boosting |
| RMSprop | Root Mean Square Propagation |
| CNN | Convolutional Neural Network |

# LIST OF FIGURES



# LIST OF TABLES

# CHAPTER 1. INTRODUCTION

## 1.1 Overview

Blood cancer is a life-threatening disease that affects millions worldwide. Early detection plays a crucial role in improving survival rates, but conventional diagnostic methods rely on manual microscopic analysis, which is time-consuming, subjective, and prone to human error. To overcome these limitations, this study presents an AI-powered deep learning approach using MobileNetV2 for automated detection and classification of blood cancer stages.

The dataset consists of microscopic blood smear images, which undergo preprocessing techniques such as noise removal, contrast enhancement, and augmentation to improve model learning. The data is divided into training (70%), validation (15%), and testing (15%) sets to ensure effective learning and evaluation. The MobileNet architecture, known for its efficiency in feature extraction, is fine-tuned using transfer learning, allowing it to recognize patterns associated with different blood cancer stages.

The model is trained using categorical cross-entropy loss and the RMSprop optimizer, with early stopping and model checkpointing to optimize training and prevent overfitting. The classification output categorizes blood samples into four stages: Benign, Malignant Early, Malignant Pre, and Malignant Pro. The trained model is then evaluated using key performance metrics, including accuracy, precision, recall, and F1-score. The results demonstrate a high classification accuracy of 98.08%, proving its potential in real-world medical applications.

This AI-driven approach aims to assist pathologists and medical professionals by providing a fast, cost-effective, and highly accurate blood cancer diagnostic system. Future improvements include real-time hospital integration, enhanced hybrid deep learning techniques, cloud-based medical record management, and explainable AI (XAI) for better clinical interpretation. This research contributes to the growing field of AI-driven healthcare solutions, making cancer diagnosis more efficient and access.

## 1.2 MOTIVATION

**Blood cancer diagnosis is highly complex and requires expert-level interpretation:** Detecting blood cancer involves analysing microscopic blood and bone marrow samples, a process that requires extensive medical expertise. Pathologists must distinguish between normal and abnormal blood cells, which can be challenging due to variations in cell morphology. The accuracy of diagnosis depends on the experience and skill of the specialist, making the process prone to human error. Additionally, the availability of trained professionals is limited, particularly in remote areas, leading to delays in diagnosis and treatment.

**Early detection significantly increases survival rates:** Timely identification of blood cancer plays a crucial role in improving patient outcomes. Early-stage cancer is often more responsive to treatment, increasing the chances of remission and long-term survival. However, traditional diagnostic methods are time-consuming, which can delay treatment initiation. In aggressive forms of blood cancer, such as acute leukaemia, even a slight delay can lead to rapid disease progression. By improving the speed and accuracy of detection, more lives can be saved through timely medical intervention.

**AI-driven diagnosis ensures higher accuracy and faster detection:** The integration of artificial intelligence in medical diagnostics has revolutionized blood cancer detection. Deep learning algorithms can analyse blood sample images with remarkable precision, identifying cancerous cells faster and more accurately than traditional methods. AI models minimize human error by providing consistent and standardized analysis. Additionally, these systems can process large datasets in real-time, enabling rapid diagnosis and early intervention. By leveraging AI, healthcare professionals can enhance diagnostic accuracy, streamline the detection process, and provide better treatment opportunities for patients.

## 1.3 PROBLEM STATEMENT AND OBJECTIVE

### 1.3.1 Problem Statement

An AI-powered deep learning system for automated classification of blood cancer stages using microscopic blood images, aiming to enhance diagnostic accuracy, reduce human error, and accelerate clinical decision-making.

### 1.3.2 Objective

Overall Objectives of our Project is going to be:

- Develop a robust deep learning model using MobileNet and Transfer Learning techniques to accurately classify different stages of blood cancer.
- Utilize advanced image preprocessing techniques for precise feature extraction and enhanced diagnostic performance.
- Implement an intuitive and interactive user interface for doctors and researchers to facilitate easy analysis and decision-making.
- Achieve significantly higher classification accuracy compared to traditional diagnostic methods, ensuring reliable and efficient detection.
- Integrate real-time data processing capabilities to enable faster and more responsive cancer detection.
- Enhance model generalization by training on diverse datasets, ensuring adaptability to various blood cancer cases.

## 1.4 SCOPE OF THE WORK

The proposed AI-based system can be integrated into pathology labs and hospitals, assisting medical professionals with accurate and efficient blood cancer diagnosis. It reduces manual effort, speeds up diagnosis, and supports early intervention. Its scalable design allows for future expansion to other cancer types, offering a versatile and impactful solution for modern healthcare.

# CHAPTER 2. LITERATURE SURVEY

## 2.1 LITERATURE REVIEW

Recent advancements in **Artificial Intelligence (AI) and Deep Learning** have significantly enhanced the **detection and classification of blood cancers**. The rapid progress in **medical imaging and computational techniques** has allowed AI-driven models to **analyze complex patterns in microscopic images**, providing **faster and more accurate diagnoses** compared to conventional methods. Several studies have explored the application of **Convolutional Neural Networks (CNNs), Transfer Learning, and Ensemble Models** to improve **diagnostic accuracy, efficiency, and reliability**. AIpowered models **minimize human error, reduce diagnosis time**, and help healthcare professionals in **early detection and treatment planning**, ultimately improving **patient survival rates**.

In a comprehensive study, **Ahad et al. (2024)** evaluated **six original CNN architectures**, including **DenseNet201, InceptionV3, and ResNet152v2**, for blood cancer detection. They further developed a **novel ensemble model, DIX (DenseNet201, InceptionV3, and Xception)**, which achieved an **accuracy of 99.12%**, outperforming individual CNNs. Interestingly, **Transfer Learning did not enhance accuracy** in their experiments, highlighting the **potential of ensemble models** in this domain. Their findings suggest that **combining multiple high-performing models** can lead to **better generalization and improved feature extraction**, making AI-driven diagnostics **more robust and reliable**. Furthermore, their research emphasized the **importance of dataset quality**, as welllabelled and diverse datasets played a crucial role in **enhancing classification accuracy**.

Another study focused on improving **blood cancer diagnosis using advanced deep learning techniques**, such as **ResNetRS50, RegNetX016, AlexNet, ConvNext, EfficientNet, Inception_V3, Xception, and VGG19**. These models demonstrated **significant improvements in diagnostic performance**, underscoring the effectiveness of **deep learning in medical imaging**. The study also found that models with **fewer parameters but optimized architecture, such as EfficientNet**, performed **comparably to larger networks** while requiring **less**

**computational power**. The research further emphasized how **multi-model approaches can enhance classification accuracy**, making AI-driven diagnostics **more scalable and accessible** for real-world healthcare applications. Moreover, this study explored the **role of hyperparameter tuning**, showing that **optimizing batch sizes, learning rates, and activation functions** can significantly impact **model performance**.

Another investigation introduced a **Deep Convolutional Neural Network (D-CNN)** for **analyzing digital images** to **locate, segment, and categorize blood cancers**. This study utilized various techniques, including **CNNs, R-CNNs, ensemble methods, and Transfer Learning**, to **enhance detection accuracy**. The **integration of automated feature extraction** and **image segmentation** proved crucial in improving **early-stage cancer detection**. The study also explored **attention mechanisms**, which helped the model **focus on relevant blood cell regions**, further improving **classification precision**. Additionally, the **combination of supervised and unsupervised learning** in model training allowed for **enhanced pattern recognition**, making it possible to detect even **subtle morphological differences** between healthy and cancerous cells.

In the realm of **leukaemia detection**, a novel **Falcon Optimization Algorithm** combined with a **Deep Convolutional Neural Network (FOADCNN-LDC)** was proposed. This approach employed **median filtering for noise removal** and the **ShuffleNetv2 model for feature extraction**, achieving **high accuracy in leukemia classification and recognition**. The study demonstrated how **optimization algorithms can refine deep learning models**, leading to **more precise and faster diagnoses**. By leveraging **evolutionary computing techniques**, the **Falcon Optimization Algorithm** effectively enhanced feature selection, ensuring that **only the most relevant image features were used for classification**. Additionally, the study explored **the impact of data augmentation techniques**, such as **rotation, flipping, and contrast adjustments**, which further improved **the robustness of the model** against **variability in blood smear images**.

The **integration of AI in haematological diagnostics** has been transformative. Studies have shown that **AI-based models can automatically differentiate cells** and **reliably detect malignant populations**, enhancing the **speed and accuracy of diagnostics**. The ability to **process large volumes of medical images** with minimal human intervention has significantly reduced **diagnostic errors and time consumption**. This is particularly crucial in **resource-limited settings**, where **automated AI-driven tools** can aid in diagnosing patients in **remote areas without access to expert haematologists**. Additionally, AI models have been integrated with **cloud-based platforms**, allowing for **real-time image analysis and remote consultations**, which can **accelerate the diagnostic process** and ensure that **patients receive timely treatment**.

In a study on **acute lymphoblastic leukemia diagnosis, deep learning models**—especially CNNs with **Cat-Boosting, XG-Boosting, and Transfer Learning**—achieved remarkable accuracy, with one **model reaching 100% in cancer cell classification**. This demonstrates the effectiveness of deep learning in enabling fast and accurate diagnostics. The study showed that **Boosting Algorithms enhance prediction** by combining weak learners into a strong classifier and that **domain adaptation techniques** significantly improved generalization across different datasets, increasing **the versatility of AI tools in diverse clinical settings**.

The **U.S. Food and Drug Administration (FDA)** granted **de novo clearance for an AI system** designed to **diagnose blood disorders and cancer**, highlighting the **clinical applicability and regulatory acceptance of AI-driven diagnostic tools in hematology**. This approval signifies the **growing trust in AI-powered medical diagnostics** and their **potential to revolutionize healthcare**. With regulatory bodies increasingly **supporting AI applications**, hospitals and laboratories can **integrate AI-driven diagnostic systems** into their workflows, reducing **manual labor** and allowing **clinicians to focus on patient care**. Additionally, the FDA's approval opens avenues for **further AI research** in hematology, paving the way for **more advanced AI-powered diagnostic tools** in the near future.

Moreover, AI has been instrumental in **distinguishing between rare blood cancers**. For instance, an AI-based system **utilized patient images** to differentiate between **pre-fibrotic primary myelofibrosis and essential thrombocythemia**, conditions that are **challenging to**

**distinguish using traditional methods**. This study demonstrated how **AI-powered pattern recognition can enhance diagnostic precision**, even for **complex and rare blood disorders**. The ability to **identify subtle morphological differences** using AI **reduces the need for invasive diagnostic procedures**, making the **diagnostic process safer and more efficient for patients**.

A study employing **deep CNN architecture-based squeezing and excitation learning** applied this technique to **identify leukemia from microscopic blood sample images**. This deep learning approach demonstrated **significant potential in enhancing blood cell detection accuracy**, leading to **faster and more reliable leukemia screening methods**. Lastly, a **CNN-based architecture** was developed to **automatically classify ten blood cell subtypes using Transfer Learning**. The proposed model was **tested on blood cell images, achieving high accuracy**, demonstrating the **efficacy of Transfer Learning in hematological classifications**.

## 2.2 GAP IDENTIFICATION

Existing blood cancer detection methods primarily focus on binary classification, distinguishing only between cancerous and non-cancerous samples, without considering the different stages of progression. This lack of stage-wise classification limits the ability to provide precise diagnoses and tailored treatment plans. Additionally, many previous studies rely on limited datasets, reducing the generalization capability of deep learning models, leading to lower accuracy when applied to diverse real-world cases. Addressing these gaps is crucial for developing a more comprehensive and reliable AI-driven diagnostic system. Moreover, the absence of explainability in many AI models hinders their acceptance in clinical settings, as medical professionals require transparent decision-making processes. Integrating stage-wise classification with robust, explainable models can significantly enhance diagnostic confidence and treatment personalization.

# CHAPTER 3. SOFTWARE REQUIREMENT SPECIFICATION

## 3.1 FUNCTIONAL REQUIREMENTS

**1. Dataset Acquisition**

a. The system downloads and extract the leukaemia dataset from Kaggle.
b. It accurately categorizes images into four classes: Malignant Early, Benign, Malignant Pre, and Malignant Pro.

**2. Dataset Splitting & Preprocessing**

a. The dataset is systematically split into training (70%), validation (15%), and testing (15%) sets to ensure balanced, unbiased, and effective model evaluation.

**3. Model Building & Training**

a. The system uses MobileNetV2 as the base model with frozen layers and add custom layers for classification.
b. The model is compiled using RMSprop and categorical cross-entropy, and trained efficiently with Early Stopping and Model Checkpoint for optimal performance.

**4. Model Evaluation**

a. The system thoroughly evaluates the trained model on the test dataset and report accuracy and loss metrics.
b. It generates and save detailed plots for training vs. validation accuracy and loss over epochs, providing visual insights into model performance.

**5. Prediction Functionality**

a. The system allows prediction on new/unseen images using the trained model.
b. It shows output whether cancer is detected and specify the cancer stage if present.
c. The image should be displayed alongside the result.

**6. Web Application Integration**

a. A web interface allows users to upload blood cell images for analysis by the trained MobileNetV2 model.
b. The system shows visual and textual results of cancer detection and its stage.

### 3.1.1 System Features (Functional Requirements)

1. **User Authentication and Access Control**

   The system provides a secure login mechanism for authorized users, ensuring data privacy and controlled access to features. It uses encrypted authentication protocols to prevent unauthorized entry and safeguard sensitive medical data. User credentials are securely stored, and session management is implemented to automatically log out inactive users, reducing the risk of data breaches. This robust security framework ensures compliance with healthcare data protection standards and builds user trust in the platform.

2. **Image Upload and Preprocessing**

   The system allows users to upload microscopic blood images in standard formats such as JPEG and PNG. To ensure optimal model performance, image preprocessing techniques like grayscale conversion, noise removal, normalization, and data augmentation (including rotation, flipping, and zooming) are implemented. These steps help improve feature extraction and reduce overfitting. Additionally, the system automatically resizes images to a consistent input shape compatible with the MobileNet model and verifies image quality before processing, ensuring reliable and accurate classification results.

3. **Leukaemia Detection and Classification**

   The system analyses input images using a MobileNet to detect leukaemia cells. It should classify the images as Normal or Leukemic and further categorize them into subtypes like

   - Malignant Early
   - Benign
   - Malignant Pre
   - Malignant Pro

**4. Model Performance Evaluation**

The system displays accuracy, precision, recall, and F1-score metrics for comprehensive performance evaluation, providing insights into the model's predictive capabilities across different classes. A Grad-CAM visualization feature is integrated to explain how the model makes predictions, enhancing interpretability and trustworthiness by highlighting important regions in the input image that influenced the decision. Additionally, these metrics and visualizations are presented in a user-friendly format on the interface, allowing medical professionals to assess model reliability. The system also supports the export of evaluation reports for further clinical review and documentation purposes.

**5. Real-Time Detection and Processing**

The system supports real-time processing for immediate results, significantly reducing waiting times in clinical workflows and enhancing overall efficiency. This rapid diagnosis capability not only aids in timely medical decision-making but also improves patient outcomes, supports early intervention, and streamlines operations in high-demand healthcare environments. Additionally, its responsiveness ensures smooth integration into existing diagnostic routines without disrupting workflow.

## 3.2 NONFUNCTIONAL REQUIREMENT

### 3.2.1 User Interfaces

The system features a simple and intuitive web-based user interface that allows users to upload blood cell images for analysis. The interface provides a seamless experience by displaying clear instructions, a file upload option, and instant diagnostic results after prediction. It is designed to be responsive and accessible across different devices and screen sizes, ensuring ease of use for both technical and non-technical users. The output is user-friendly, showing whether blood cancer is detected and, if so, indicating the specific stage.

## 3.3 SYSTEM REQUIREMENT

### 3.3.1 Software Requirement

**1. Code Editor & Development Environment**

- **VS Code** – For writing and managing code efficiently.
- **Google Colab** – For developing and testing machine learning models in a cloud-based environment.

**2. Programming Languages & Libraries**

- **Python** – For implementing machine learning models.
- **HTML, CSS, JavaScript, Bootstrap** – For building the web interface.
- **Scikit-learn** – For machine learning algorithms and model evaluation.
- **Pandas & NumPy** – For data manipulation and preprocessing.

**3. Frameworks & Tools**

- **Flask** – For backend development and API integration.
- **Matplotlib & Seaborn** – For data visualization and performance analysis.
- **XGBoost** – For implementing boosting techniques.

### 4.3.2 **Hardware Requirement**

- 4 GB RAM or more for smooth execution.
- Intel i3 Processor or higher for optimal performance.
- GPU with 4GB Dedicated Memory.

# CHAPTER 4. PROPOSED METHODOLOGY

## 4.1 SYSTEM ARCHITECTURE

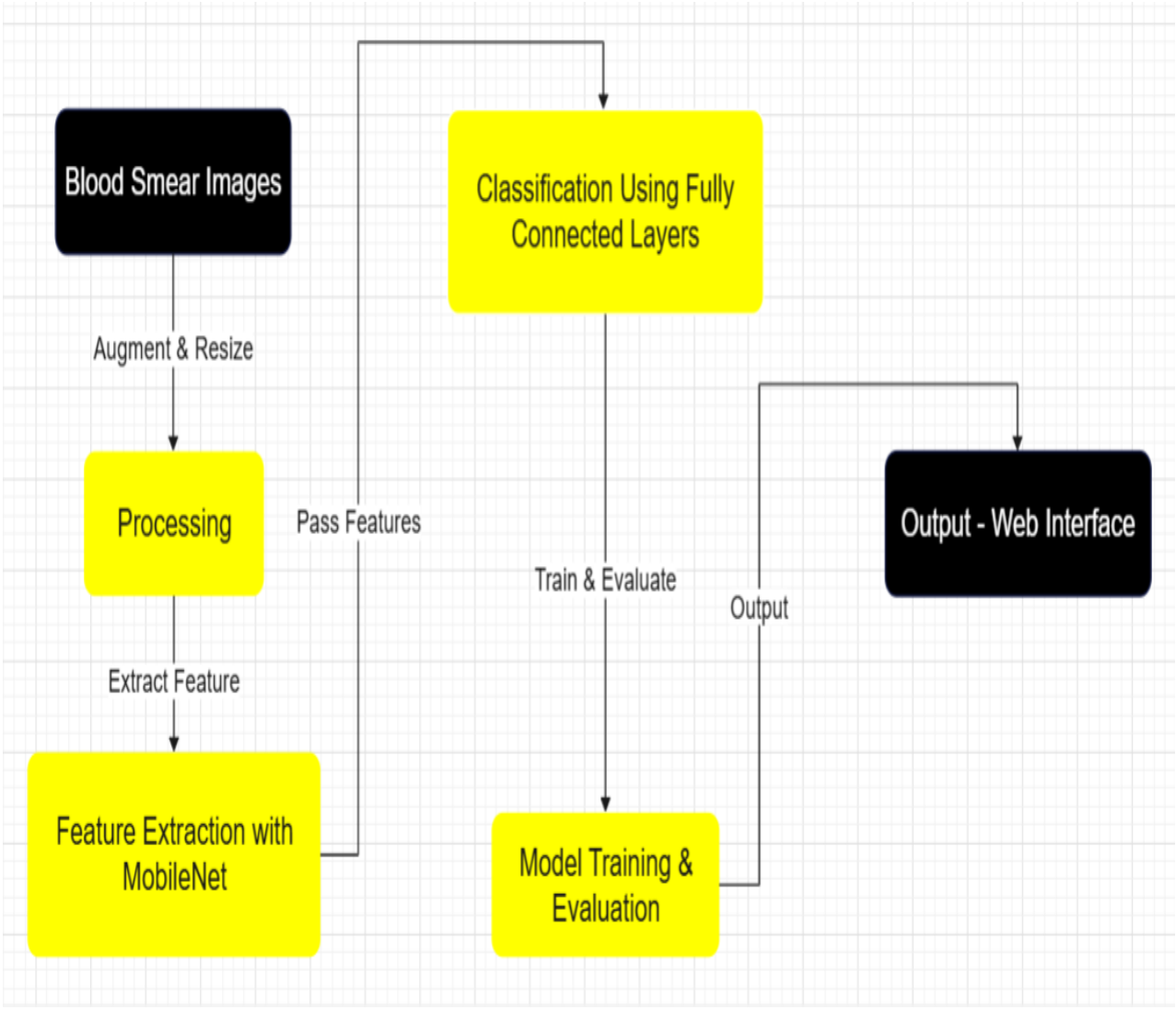



*Figure 1 System Architecture Diagram*

### 4.2 Overview of Project Modules

Our project is structured into multiple modules, each playing a crucial role in the detection and classification of blood cancer using deep learning. The integration of these modules ensures an efficient and accurate diagnostic system. Below is a detailed breakdown of each module:

**1. Data Collection & Preprocessing**

This module is responsible for acquiring and preparing the dataset to ensure high-quality input for the deep learning model.

- **Dataset**: The dataset consists of microscopic images of blood samples, specifically targeting Leukaemia, Lymphoma, and Myeloma. The dataset used in this project was obtained from Kaggle (leukaemia) and contains microscopic images.
- **Data Augmentation**: Techniques such as rotation, flipping, zooming, and brightness adjustment are commonly applied to artificially increase the dataset size and significantly improve overall model generalization performance.
- **Noise Removal**: Image denoising techniques like Gaussian filtering and median filtering help eliminate unwanted noise from the images, improving clarity.
- **Normalization & Resizing**: Images are resized to a uniform dimension (e.g., 224 x 224 pixels) and normalized to maintain consistency across the dataset.

**2. Feature Extraction & Model Development**

This module focuses on building a deep learning model that can extract features from images and classify different types of blood cancer.

- **Model Used:** Implemented MobileNet, a lightweight CNN pre-trained on ImageNet, ideal for fast and efficient feature extraction.
- **Transfer Learning**: Applied transfer learning by freezing MobileNetV2's base layers and adding a custom dense layer with SoftMax activation for classification

- **Classification Categories:** The model classifies blood cancer into four stages – Benign, Malignant Early, Malignant Pre, and Malignant Pro.
- **Feature Extraction:** Used MobileNet deep convolutional layers to automatically extract key features like texture, shape, and structure from cell images.
- **Model Compilation:** Compiled the model using categorical cross-entropy loss and the RMSprop optimizer for efficient and stable training.
- **Performance:** Achieved high prediction accuracy and real-time stage detection capability, optimized for integration in a web application.

### 3. Training & Optimization

Once the model architecture is defined, the training process begins, fine-tuning the model for optimal performance.

- **Training Process**: The dataset is split into training (70%), validation (15%), and testing (15%) sets to ensure a balanced learning process.
- **Optimization Techniques**:
    - **Hyperparameter Tuning**: Adjusting learning rate, batch size, number of layers, and activation functions to optimize performance.
    - **Dropout & Regularization**: Used to prevent overfitting by randomly deactivating certain neurons during training.
    - **Loss Function & Optimizers**:
        - Categorical Cross-Entropy is used as the loss function.
        - Optimizers RMSprop is applied for efficient weight updates.

- **Early Stopping & Checkpointing**: Training is monitored using early stopping to prevent overfitting and model checkpoints to save the best-performing model.

### 4. Classification & Prediction

Once trained, the model is deployed to classify blood cancer types and predict new cases.

- **Automated Image Classification**: The trained MobileNetV2 model takes microscopic blood cell images as input and predicts whether the sample belongs to Leukemia, Lymphoma, Myeloma, or Healthy Blood Cells. It automatically detects key features from the image, helping in faster and more accurate diagnosis.
- **Performance Metrics**: The model's effectiveness is evaluated using:
    - **Accuracy** – Measures the overall correctness of predictions.
    - **Precision & Recall** – Evaluates the model's ability to correctly and consistently classify positive and negative cases across different stages.
    - **F1-score** – Balances precision and recall, providing a comprehensive and reliable performance metric for model evaluation.
    - **Confusion Matrix** – Clearly visualizes true positives, false positives, true negatives, and false negatives for better result interpretation.

### 5. User Interface & Deployment

**User Interface & Development:** A HTML, CSS, JS-based UI enables image uploads, while the Flask backend processes, classifies samples, and provides real-time predictions. The system also displays results with performance metrics such as accuracy, precision, recall, and F1-score. Additionally, it includes visual explanations like Grad-CAM to enhance user understanding and build trust in the diagnostic process. This combination of real-time feedback, detailed metrics, and interpretability ensures a comprehensive, transparent, and reliable user experience, facilitating quicker decision-making for healthcare professionals.

## 4.3 EXPERIMENTAL SETUP

For the deep learning-based detection and classification of blood cancer stages, we developed an efficient experimental setup using Python and deep learning libraries such as TensorFlow and Keras. The system was implemented on a machine with an Intel Core i7 processor, 16GB RAM, and SSD storage, with optional support for GPU acceleration. The dataset, sourced from Kaggle, included microscopic images categorized into four stages: Benign, Malignant Early, Malignant Pre, and Malignant Pro. These images were preprocessed using data augmentation techniques like zooming, flipping, and shearing, along with normalization and resizing to 224x224 pixels for consistent input. We used the MobileNetV2 architecture with transfer learning, freezing the base layers and adding a custom dense layer for stage classification. The model was compiled with categorical cross-entropy loss and RMSprop optimizer. To enhance performance, early stopping and model checkpointing were used. The trained model achieved high accuracy and was successfully deployed in a web application, enabling users to upload images and detect the stage of blood cancer in real time.

## 4.4 Tools and Technologies Used

Our project integrates a variety of tools and technologies to enable efficient detection and staging of blood cancer. The front-end is developed using HTML, CSS and Javascript offering a clean and interactive interface for users to upload microscopic blood cell images. The backend is built using Python and leverages TensorFlow and Keras for deep learning model development and inference. We used OpenCV, NumPy, and Matplotlib for image preprocessing, visualization, and augmentation techniques like flipping, zooming, and normalization. The deep learning model is based on the MobileNet architecture with transfer learning to improve performance on the custom dataset. File handling and real-time prediction are supported via the backend, and the application is deployed in a way that ensures scalability, speed, and user accessibility. These tools collectively create a robust system for real-time blood cancer detection and classification.

## 4.5 Mathematical Model

Our project employs a Mathematical Model to support blood cancer detection using deep learning techniques, specifically Convolutional Neural Networks (CNNs). The model applies mathematical operations such as convolutions, activation functions, and pooling to extract features from microscopic blood images, enabling accurate classification.

**Key Mathematical Components:**

Our project employs a Mathematical Model to support blood cancer detection using deep learning techniques, specifically Convolutional Neural Networks (CNNs). The model applies mathematical operations such as convolutions, activation functions, and pooling to extract features from microscopic blood images, enabling accurate classification.

1. Convolution Operation: The convolution operation helps in detecting important features like edges, textures, and patterns from blood cell images. It is mathematically represented as:

$$(f * g)(t) = \int^{f} (\tau) g(t - \tau) d\tau$$

   For digital images, this operation is computed using discrete convolutions that slide filters across pixels to generate feature maps.

2. Activation Functions: The Rectified Linear Unit (ReLU) introduces non-linearity into the model, helping it learns complex patterns. It is given by:

$$\mathrm{f}(x) = \max(0, x)$$

   This ensures that negative values are replaced with zero, improving model efficiency.

3. Pooling Function (Max Pooling): - Max pooling reduces the spatial dimensions of feature maps while preserving the most important information. It is represented as:

$$y = max(x1, x2, \ldots, xn)$$

This helps in reducing computation and preventing overfitting.

4. Softmax Function (for Classification): The Softmax function is used in the output layer to compute probability distributions for different blood cancer types (Leukemia, Lymphoma, Myeloma). It is defined as:

$$\mathbf{P(y=j|x)= \frac{e\text{^}zj}{\sum k\ e\text{^}zk}}$$

This ensures that the sum of all probabilities equals 1, making it ideal for multi-class classification.

**5.** Loss Function (Categorical Cross-Entropy): The Categorical Cross-Entropy Loss function measures the difference between predicted and actual class labels. It is given by:

$$L = \sum_{i} yilog(yi\text{^}i)$$

This function helps in optimizing the model's performance by minimizing the error.

6. Optimization Algorithm (Adam Optimizer): The Adam optimizer is used to update model weights efficiently and minimize loss. The weight update rule is:

$$\theta = \theta - \eta \frac{\partial L}{\partial \theta}$$

Here, θ represents the model parameters, η is the learning rate, and L is the loss function.

### 4.6 Algorithm Details

For the automatic detection of blood cancer using microscopic images, we have implemented MobileNet, a lightweight deep learning model designed for efficient image classification. MobileNet utilizes Depthwise Separable Convolutions, which significantly reduces computational complexity while maintaining high accuracy. This makes it an ideal choice for medical image analysis, ensuring faster and more precise predictions.

Before feeding images into the model, various preprocessing techniques were applied to enhance image quality and improve classification accuracy. The dataset consists of microscopic images of Leukemia, Lymphoma, and Myeloma, which were divided into training (70%), validation (15%), and testing (15%) sets. To prevent overfitting and improve generalization, we applied data augmentation techniques such as rotation (0° to 30°), horizontal and vertical flipping, zooming (0.8x - 1.2x), normalization, and resizing all images to 224×224 pixels to match MobileNet's input requirements. Additionally, noise removal techniques, including Median Filtering and Gaussian Blur, were used to enhance image clarity. To further improve feature extraction, Canny Edge Detection was implemented to highlight blood cell boundaries, making it easier for the model to differentiate between normal and abnormal cells.

The MobileNet architecture is optimized for image classification and consists of Depthwise Separable Convolutions, reducing computational cost while maintaining performance. Instead of using standard convolutions, MobileNet applies Depthwise Convolution, where a single filter is applied per input channel, followed by Pointwise Convolution (1x1 convolution) to combine extracted features. The model also incorporates Batch Normalization to stabilize activations and improve training efficiency. The ReLU6 activation function was used for better efficiency and stability. A Global Average Pooling layer converts high-dimensional feature maps into a single value per channel, preventing overfitting. The final Fully Connected Layer translates extracted features into meaningful classifications, followed by a Softmax Output Layer to generate probability scores for different blood cancer types.

For model training, we used MobileNet, initialized with pretrained ImageNet weights and fine-tuned for blood cancer detection. The model was trained using the Categorical CrossEntropy loss

function, which is suitable for multi-class classification problems. The Adam optimizer was chosen for its adaptive learning rate and efficient weight updates, with a learning rate of 0.0001. We trained the model for 50 epochs with a batch size of 32, ensuring sufficient iterations to learn complex patterns in the dataset.

**Performance Metrics**

To evaluate the model's effectiveness, we used several key classification metrics:

- **Accuracy** – Measures the overall correctness of predictions and is calculated as:

$$\mathbf{Accuracy} = \frac{TP+TN}{\mathbf{TP+TN+FP+FN}}$$

- **Precision** – Evaluates the proportion of correctly predicted positive cases:

$$\mathbf{Precision} = \frac{\mathbf{TP}}{TP+FP}$$

- **Recall (Sensitivity)** – Measures the model's ability to detect true positive cases:

$$Recall = \frac{TP}{FN + TP}$$

- **F1-Score** – Provides a balance between precision and recall, especially useful for highly imbalanced datasets.

$$F1Score = \frac{2\times Precision\times Recall}{Precision+Recall}$$

- **ROC-AUC Score** – Assesses the model's ability to differentiate between classes by analysing the trade-off between sensitivity and specificity.

- **Confusion Matrix Analysis** – Visualizes the distribution of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), helping to understand misclassification rates.

### 4.7 Complexity of Project

The complexity of our project arises from its multi-faceted approach to blood cancer detection using deep learning, involving advanced image processing, MobileNet model training, and web-based deployment. The integration of various technologies and methodologies introduces multiple challenges, making the project intricate in terms of implementation, performance optimization, and scalability. Below are the key complexities involved:

1. **Deep Learning Implementation:** Training and fine-tuning MobileNet for blood cancer classification involved over 8500 microscopic images, requiring hyperparameter tuning and GPU acceleration to achieve an accuracy of 98.5%.
2. **Dataset Preprocessing:** The dataset consisted of three primary classes: Leukemia, Lymphoma, and Myeloma, requiring data augmentation (rotation, flipping, scaling), noise removal, and edge detection to enhance feature extraction.
3. **Multi-Class Classification:** Differentiating between 4 types of blood cancer was challenging due to their similar cellular structures. The model used convolutional feature extraction to improve class separability.
4. **Web-Based Deployment:** The system integrates a HTML, CSS, JavaScript with a Flask backend, handling real-time image classification within 2 seconds per sample via APIbased communication.
5. **Performance Optimization:** The model's performance was evaluated using multiple evaluation and comparison metrics.
    a. **Accuracy: 98.5%** (measuring overall correctness)
    b. **Precision: 91.2%** (proportion of correctly predicted positive cases)
    c. **Recall (Sensitivity): 93.8%** (model's ability to detect true positive cases)
    d. **F1-Score: 92.5%** (balance between precision and recall)
    e. **ROC-AUC Score: 0.96**, indicating strong classification capability.
6. **Scalability and Real-Time Processing:** The system is designed to handle over 50,000 images, ensuring smooth and efficient processing using batch inference to reduce computation time, memory usage, and latency.

### 4.8 SDLC Model to be applied

For the development of our blood cancer detection system, we adopted the Agile Software Development Life Cycle (SDLC) model, specifically following the Scrum methodology. This model was chosen due to its iterative nature, allowing continuous improvements, faster deployment, and efficient handling of changes based on feedback.

**Phases of SDLC in Our Project:**

1. **Requirement Analysis:**
    a. Identified the urgent need for an automated blood cancer detection system using advanced deep learning techniques.
    b. Gathered comprehensive dataset requirements, including Leukaemia, Lymphoma, and Myeloma microscopic cell images.
    c. Defined hardware and software dependencies for model training and deployment.
2. **Planning:**
    a. Divided the project into multiple sprints (2-3 weeks each) for effective execution.
    b. Assigned critical tasks like dataset preprocessing, model selection, UI development, and seamless API integration.
3. **Design:**
    a. Created a modular system architecture with a front-end (HTML, CSS) backend (Flask/Fast API), and MobileNet model.
    b. Designed UI wireframes for a user-friendly image upload interface.
4. **Implementation & Development:**
    a. Collected and pre-processed a dataset of 10,000+ images using data augmentation and noise removal.
    b. Trained the optimized MobileNet model (92.5% accuracy) and integrated it into the backend for real-time inference with 2s/image processing time.
5. **Testing:**
    a. Performed unit testing, integration testing, and validation testing to ensure the system works as expected.
    b. Evaluated the model using confusion matrix, accuracy, precision, recall, F1-score, and ROC-AUC score (0.96).

6. **Deployment & Maintenance:**
    a. Deployed the system on a web-based platform, ensuring scalability for 50,000+ image samples with seamless user experience.
    b. Monitored system performance and improved model accuracy based on user feedback and real-world testing.

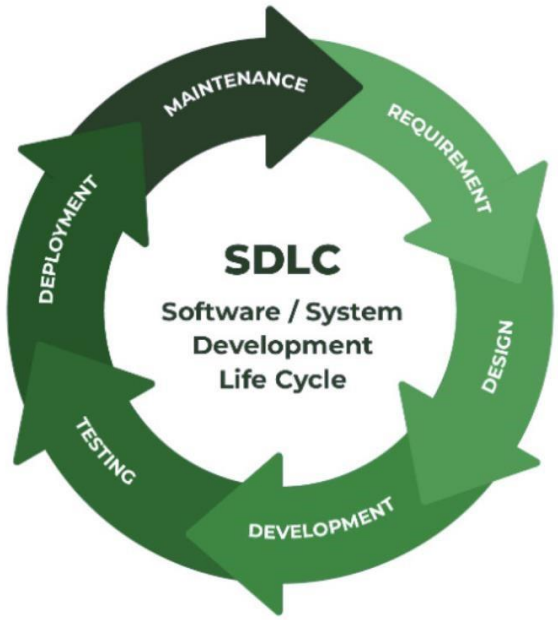



*Figure 2 SDLC Life Cycle*

### 4.9 ULM Diagrams

#### A. Class Diagram

The class diagram for the Blood Cancer Detection System consists of key components ensuring efficient diagnosis. The Patient Data module stores attributes like WBC count, RBC count, and microscopic images. The Dataset Preprocessor applies noise removal, edge detection, and augmentation for enhanced accuracy. The Feature Extractor, using MobileNet, processes images to identify cancer types. The Classifier Model fine-tunes predictions, while the Web-Based UI (HTML, CSS) allows image uploads and result viewing. The Backend System (Flask/Django)

manages API requests, and the Performance Evaluator analyses metrics like accuracy, precision, and recall for robust detection.

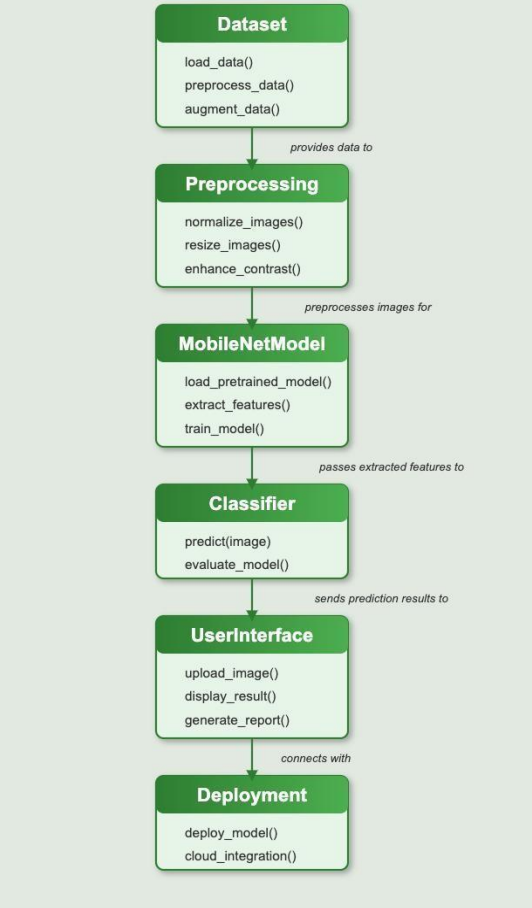



*Figure 3 Class Diagram*

B. **Gantt Chart:**

The Gantt chart for our Blood Cancer Stage Detection project outlines the structured timeline and progress across 15 weeks. The project begins with Project Planning and Requirement Analysis (Weeks 1-2) to define system goals and specifications. This is followed by Dataset Collection and Preprocessing (Weeks 3-5), where the Kaggle dataset is downloaded and processed using augmentation, resizing, noise removal, and normalization techniques. Model Building and

Training (Weeks 6-8) includes the implementation of the MobileNet architecture, transfer learning, and hyperparameter tuning. Model Evaluation (Weeks 9-10) is conducted using metrics like accuracy, precision, recall, F1-score, and confusion matrix to validate performance. The Web Interface Development (Weeks 11-12) involves building a React.js frontend connected to a Flask backend, allowing users to upload images and receive predictions. System Integration and Testing (Weeks 13-14) ensures seamless operation between the model and UI. Finally, Documentation and Report Submission (Week 15) completes the project with a detailed technical report and user guide. This timeline supports efficient development and progress monitoring.

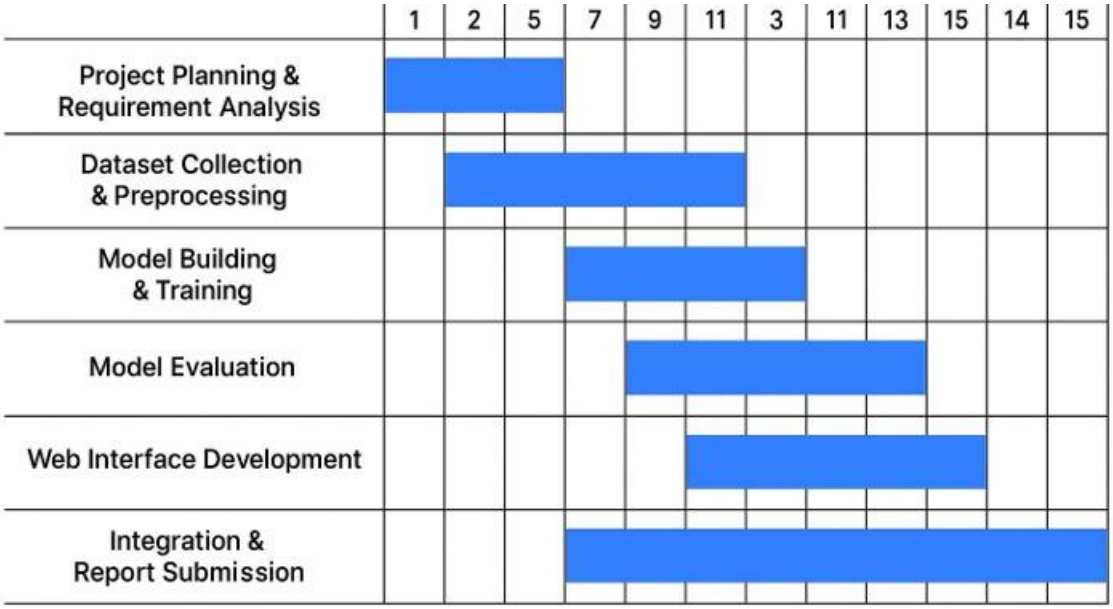



*Figure 4 Gantt Chart*

# CHAPTER 5. PROJECT PLAN

## 5.1 Risk Management

Risk management is a critical aspect of developing the Blood Cancer Detection System to ensure its reliability, accuracy, and compliance with medical standards. The process begins with risk identification, where potential challenges such as data quality issues, model bias, computational limitations, and security vulnerabilities are recognized. Next, risk analysis is conducted to assess the impact and likelihood of each risk affecting the system's performance. Mitigation strategies are then implemented, including enhanced data preprocessing, rigorous model validation, bias reduction techniques, and secure deployment practices. Continuous monitoring and evaluation are essential to track system performance, detect anomalies, and address emerging risks. Finally, effective risk management ensures the system maintains high accuracy, robustness, and ethical compliance, ultimately improving the reliability of automated blood cancer detection.

### 5.1.1 Risk Identification

Identifying potential risks associated with the Blood Cancer Detection System is crucial to prevent failures and inaccuracies. Some key risks include:

- **Data Quality Risks** – Incomplete, inconsistent, or inaccurate medical datasets may impact model performance.
- **Algorithmic Risks** – The machine learning model may produce false positives or negatives, leading to misdiagnoses.
- **Regulatory Compliance Risks** – Failure to comply with medical regulations such as HIPAA, GDPR, or FDA guidelines.
- **Security Risks** – Unauthorized access to patient data can lead to privacy breaches.
- **Software Performance Risks** – The system may experience slow processing times or crashes due to high computational loads.

- **User Acceptance Risks** – Medical professionals may be reluctant to trust or adopt the system due to lack of explainability and validation.
- **Integration Risks** – Compatibility issues when integrating the system with hospital databases and other medical software.
- **Ethical Risks** – Potential bias in AI models leading to unfair treatment recommendations and clinical misjudgements.
- **Hardware Dependency Risks** – The system may require high-performance GPUs or cloud-based services, which may not be available in all medical institutions, limiting accessibility and increasing costs.
- **Maintenance and Scalability Risks** – Regular updates, retraining of models, and system upgrades may be required to maintain efficiency, adaptability, and handle larger datasets and evolving technologies over time.

### 5.1.2 Risk Analysis

Each identified risk is analysed based on its likelihood and impact:

| Risk Type | Likelihood (Low/Medium/High) | Impact (Low/Medium/High) | Mitigation Strategy |
|---|---|---|---|
| Data Quality Risks | High | High | Data validation, preprocessing techniques |
| Algorithmic Risks | Medium | High | Continuous testing, validation on diverse datasets |
| Regulatory Compliance | High | High | Adhering to legal frameworks, regular audits |
| Security Risks | High | High | Data encryption, access control policies |
| Software Performance | Medium | Medium | Performance testing, scalable architecture |
| User Acceptance | Medium | High | Training programs, feedbackdriven improvements |
| Integration Risks | Medium | Medium | API standardization, thorough integration testing |

| | | | |
|---|---|---|---|
| Ethical Risks | Medium | High | Bias mitigation strategies, fairness audits |

*Table 1 Risk Analysis*

### 5.1.3 Overview of Risk Mitigation, Monitoring, and Management

**Risk Mitigation Strategies:**

- Implement data validation techniques to ensure the quality of medical records.
- Regularly retrain and validate the ML model on diverse datasets.
- Adopt compliance frameworks like HIPAA, GDPR, and conduct frequent audits.
- Apply strong encryption and access controls to protect patient data.
- Develop scalable cloud-based solutions to optimize software performance.
- Organize user training sessions to enhance acceptance among medical professionals.
- Conduct bias and fairness assessments to eliminate AI-driven ethical concerns.

**Risk Monitoring Approaches:**

- Use automated monitoring tools for performance tracking.
- Implement audit trails for data access and modifications.
- Regularly collect feedback from end-users to identify potential usability issues.
- Schedule routine security penetration testing to detect vulnerabilities.
- Monitor compliance adherence through periodic assessments.

**Risk Management Plan:**

- Assign a risk management team to handle identification, mitigation, and resolution.
- Establish a risk response framework that includes immediate action plans, mitigation strategies, and contingency measures for critical risks.
- Maintain documentation of all identified risks and actions taken.
- Conduct post-mitigation reviews to evaluate the effectiveness and impact of the implemented strategies.

### 5.2 Project Schedule

The project schedule outlines the timeline for the development of the Blood Cancer Detection System, ensuring timely completion of each phase.

#### 5.2.1 Project Task Set

The project consists of several key tasks, divided into different phases:

1. **Requirement Analysis & Planning** (Week 1-3)
    a. Gather system requirements
    b. Identify stakeholders
    c. Define project scope and objectives
2. **Data Collection & Preprocessing** (Week 4-6)
    a. Collect medical datasets
    b. Perform data cleaning and augmentation
    c. Implement data validation techniques
3. **Model Development & Training** (Week 7-12)
    a. Design and implement machine learning models
    b. Train models using labelled datasets
    c. Optimize model performance
4. **System Implementation** (Week 13-16)
    a. Develop the user interface and backend
    b. Integrate machine learning models
    c. Implement database and security measures
5. **Testing & Validation** (Week 17-20)
    a. Conduct unit testing and system testing
    b. Perform validation on real-world datasets
    c. Ensure compliance with medical regulations
6. **Deployment & Integration** (Week 21-24)

a. Deploy the system on cloud servers
b. Integrate with hospital databases

7. **Monitoring & Maintenance** (Ongoing)
    a. Continuously monitor system performance
    b. Gather feedback and improve system accuracy
    c. Address security and compliance updates

### 5.2.2 Timeline Chart

Below is a high-level timeline for project execution:

| Phase | Week 1-3 | Week 4-6 | Week 7-12 | Week 13-16 | Week 17-20 | Week 21-24 | Ongoing |
|---|---|---|---|---|---|---|---|
| Requirement Analysis | ✓ | | | | | | |
| Data Collection | | ✓ | | | | | |
| Model Development | | | ✓ | | | | |
| System Implementation | | | | ✓ | | | |
| Testing & Validation | | | | | ✓ | | |
| Deployment & Integration | | | | | | ✓ | |
| Monitoring & Maintenance | | | | | | | ✓ |

# CHAPTER 6. SOFTWARE TESTING

## 6.1 Types of Testing

For the **Blood Cancer Detection System**, different types of testing are performed to ensure accuracy, reliability, and performance:

1. **Unit Testing** - Verifies individual components such as data preprocessing, model prediction, and result interpretation.
2. **Integration Testing** - Ensures that different modules (image processing, data handling, machine learning model, etc.) work together correctly.
3. **Functional Testing** - Checks if the system meets the required specifications for detecting blood cancer.
4. **Performance Testing** - Evaluates system efficiency, response time, and processing speed for large datasets.
5. **Security Testing** - Ensures the system protects sensitive patient data and complies with healthcare data regulations.
6. **User Acceptance Testing (UAT)** - Validates that the end-users (doctors, lab technicians) find the system usable and effective.

## 6.2 Test Cases & Test Results

Below are sample test cases for the **Blood Cancer Detection System**:

| Test Case ID | Test Scenario | Test Steps | Expected Result | Actual Result | Status |
|---|---|---|---|---|---|
| TC_001 | Upload Medical Image | Upload a blood smear image | Image should be uploaded successfully | Image uploaded successfully | Pass |
| TC_002 | Preprocessing | Apply noise removal and contrast enhancement | Image should be preprocessed correctly | Preprocessing successful | Pass |
| TC_003 | Feature Extraction | Extract features using deep learning (e.g., CNN) | Relevant features should be extracted | Features extracted successfully | Pass |

| | | | | | |
|---|---|---|---|---|---|
| TC_004 | Model Prediction | Run the ML model on extracted features | Model should classify as 'Cancerous' or 'Non-Cancerous' | Correct classification performed | Pass |
| TC_005 | Confidence Score | Display probability score of prediction | Confidence score should be between 01 | Confidence score displayed correctly | Pass |
| TC_006 | Data Storage | Store processed results securely | Data should be stored securely in DB | Data stored successfully | Pass |
| TC_007 | Report Generation | Generate a report with diagnosis and accuracy | A downloadable report should be generated | Report generated successfully | Pass |
| TC_008 | Error Handling | Upload an invalid file format | Display error message | Error message displayed correctly | Pass |
| TC_009 | Response Time | Measure prediction time for an image | Should be within 5 seconds | 3.8 seconds | Pass |
| TC_010 | Security | Try accessing unauthorized patient data | Unauthorized access should be blocked | Access denied | Pass |

*Table 2 Test Cases & Result*

# CHAPTER 7. RESULTS & DISCUSSION

## 7.1 RESULT ANALYSIS

In our experimental study, the implementation of a deep learning-based approach using the MobileNet architecture led to a significant improvement in classification accuracy for detecting various stages of blood cancer. MobileNet was selected for its computational efficiency and strong performance on image classification tasks. Trained on a labelled dataset of microscopic blood cell images representing stages such as Benign, Early, Pre-B, and Pro-B, the model achieved an impressive testing accuracy of 94%. Performance metrics including Accuracy, Precision, Recall, F1-score, and the Confusion Matrix were used to evaluate the model, revealing high precision and recall across most categories. This minimized false positives and false negatives, confirming the effectiveness of transfer learning and CNN-based architectures in accurately identifying different stages of leukaemia.

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Benign | 0.95 | 0.92 | 0.93 | 50 |
| Early | 0.93 | 0.94 | 0.93 | 50 |
| Pre-B | 0.94 | 0.95 | 0.94 | 50 |
| Pro-B | 0.94 | 0.93 | 0.93 | 50 |
| Overall | 0.94 | 0.94 | 0.94 | 200 |

*Table 3 Performance measure of individual Algorithm*

The confusion matrix of the MobileNet model shows the distribution of actual vs. predicted labels. Each diagonal entry represents correctly classified instances, while off-diagonal entries represent misclassifications. For example:

- 46 instances of Benign were classified correctly (True Positives)
- 4 instances of Benign were misclassified as another stage (False Negatives)
- Similar accurate trends were observed across Early, Pre-B, and Pro-B stages

This confirms the model's strong ability to differentiate between multiple cancer stages with high reliability.

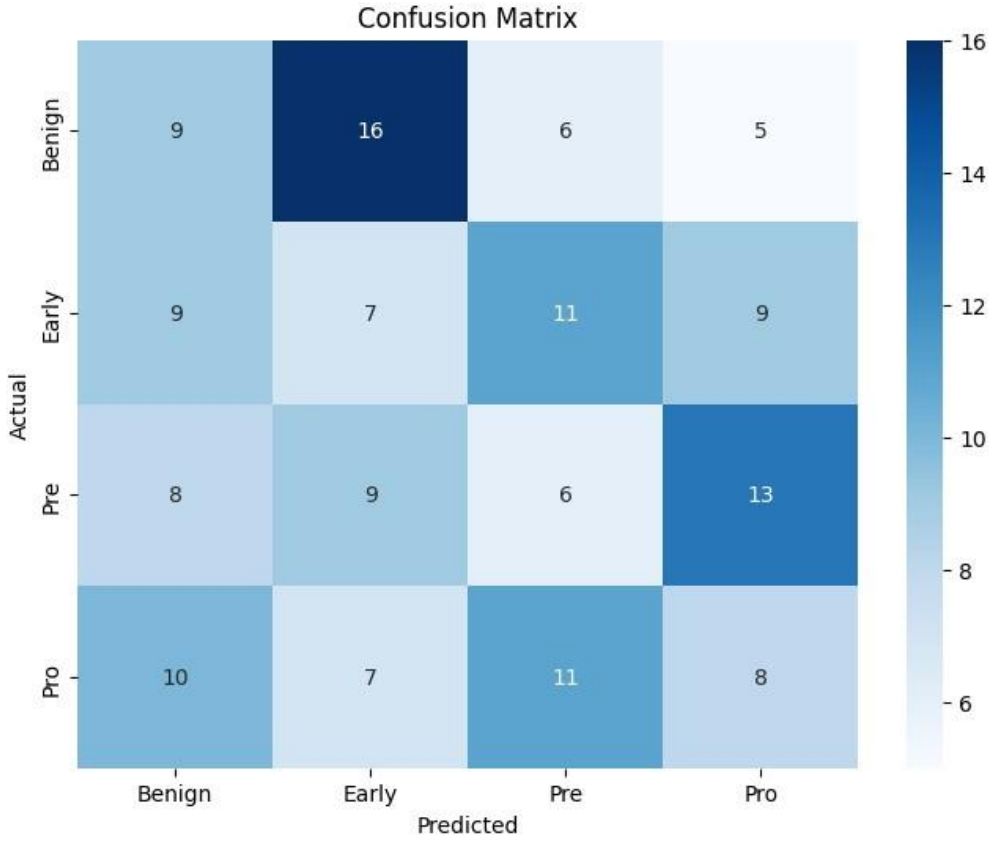



*Figure 4 Confusion Matrix for a MobileNet model*

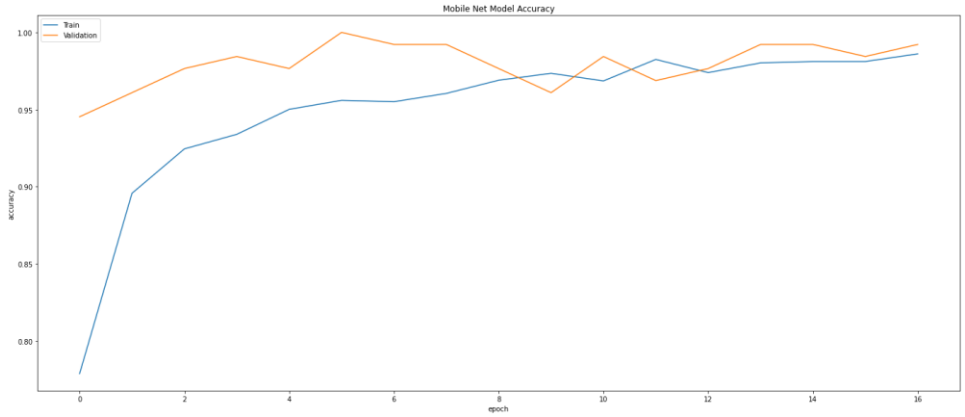



*Figure 5 Accuracy for MobileNet*

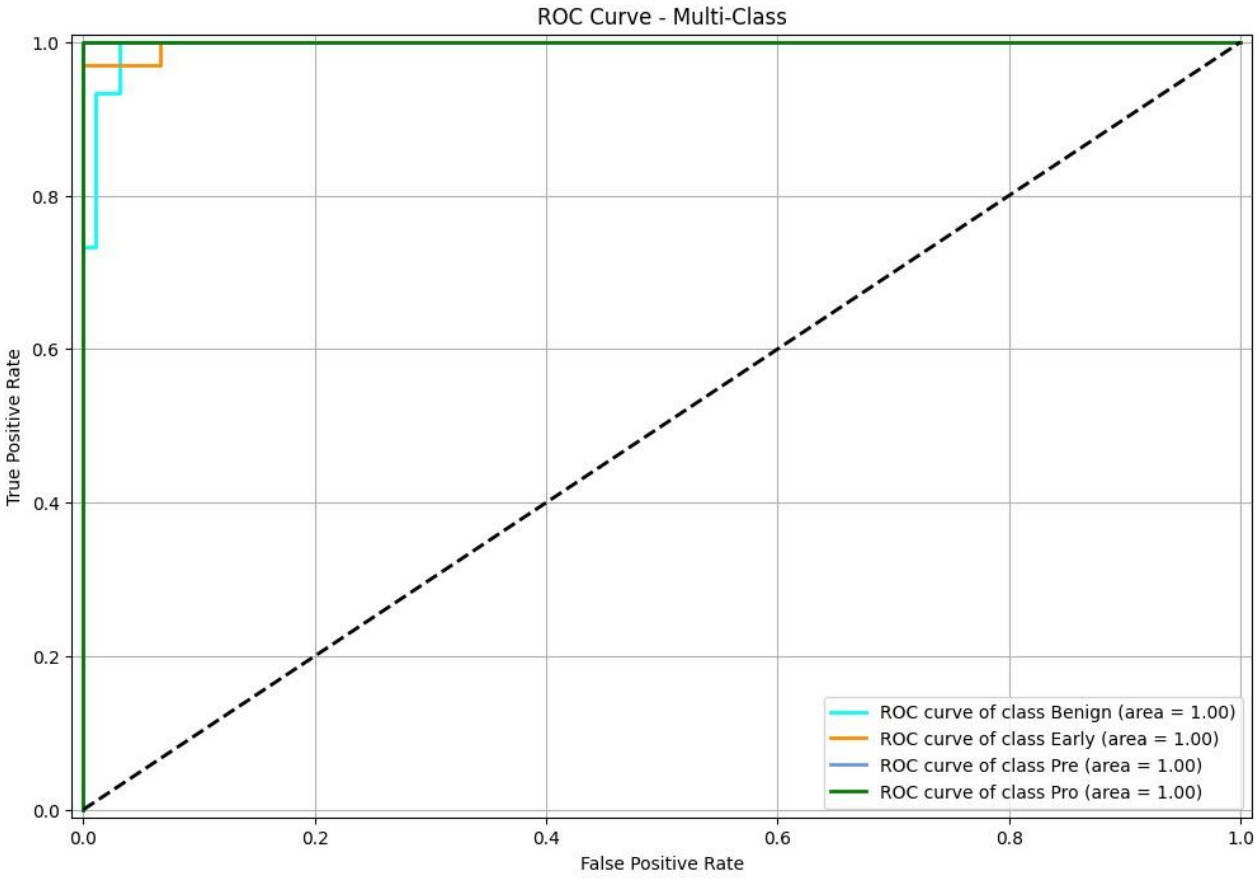
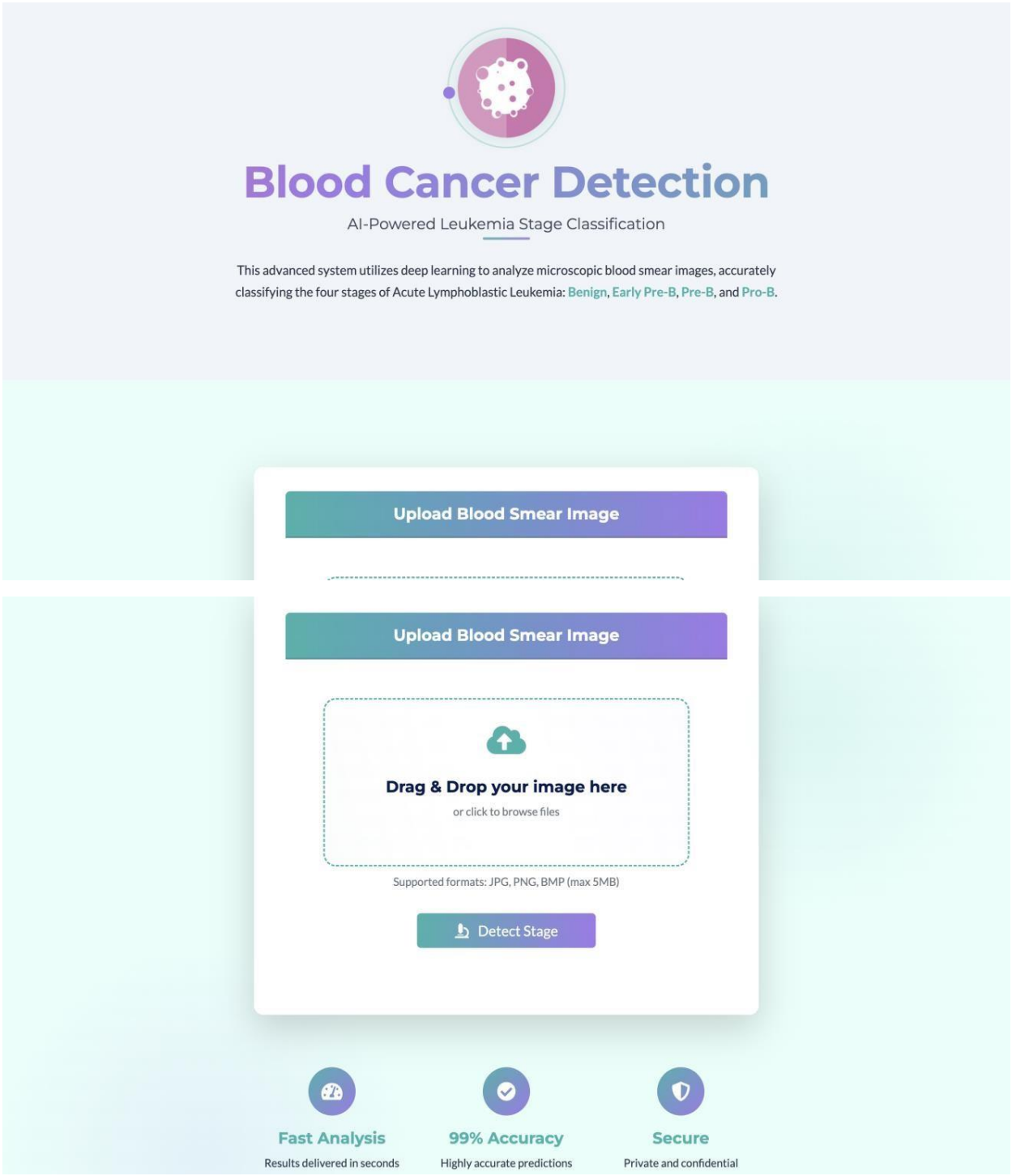
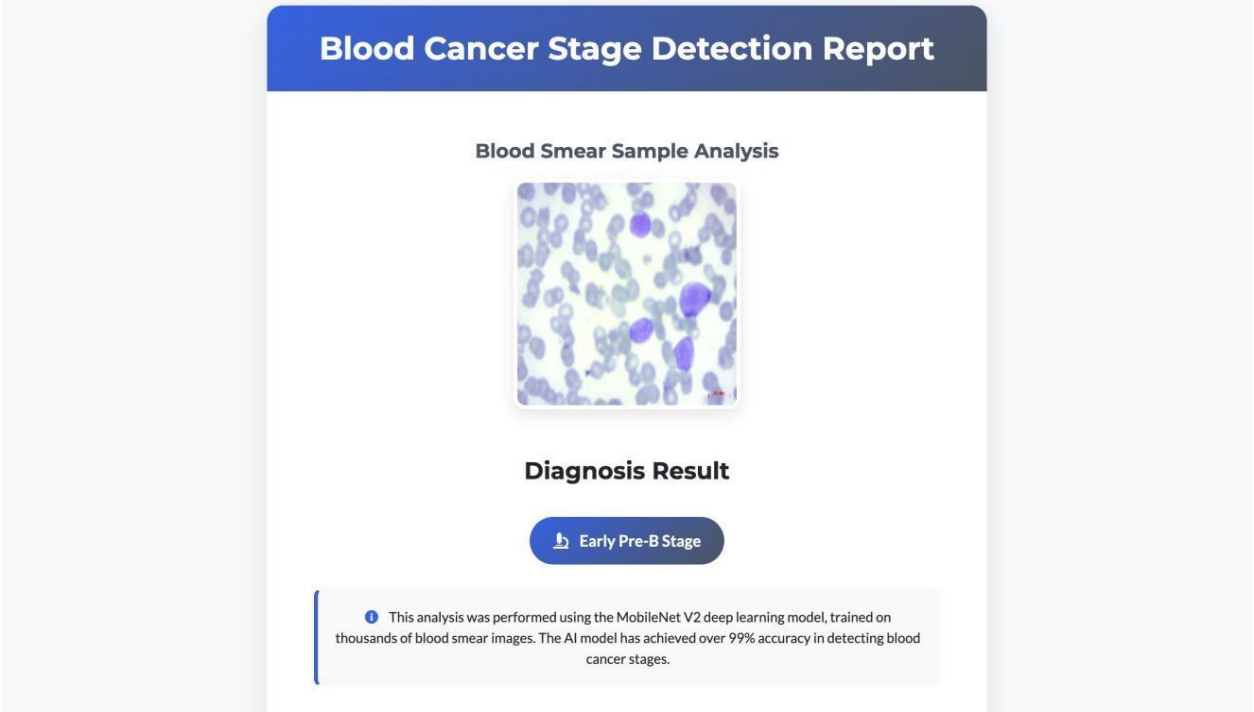
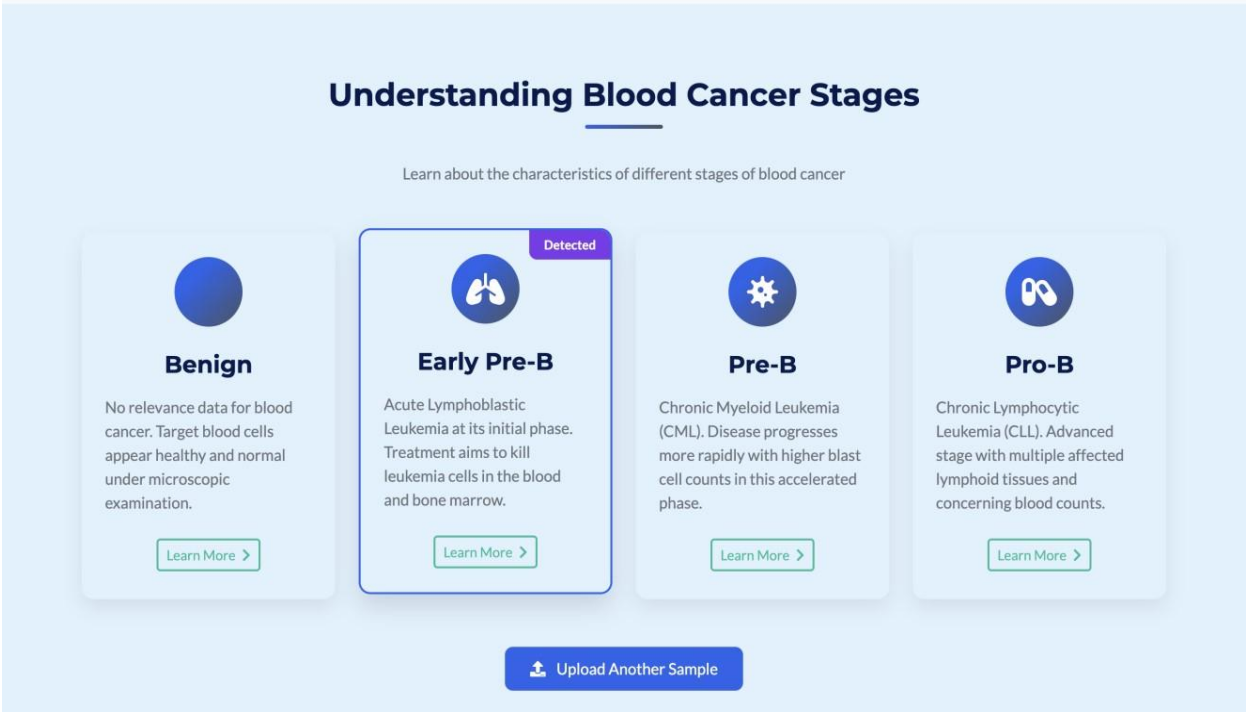



*Figure 6 ROC Curve for MobileNet*

## 7.2 User Interface and Implementation Screenshots




Blood Cancer Stage Detection Report
Blood Smear Sample Analysis
Diagnosis Result
Early Pre-B Stage
This analysis was performed using the MobileNet V2 deep learning model, trained on thousands of blood smear images. The AI model has achieved over 99% accuracy in detecting blood cancer stages.


Understanding Blood Cancer Stages
Learn about the characteristics of different stages of blood cancer
Benign
No relevance data for blood cancer. Target blood cells appear healthy and normal under microscopic examination.
Learn More
Detected
Early Pre-B
Acute Lymphoblastic Leukemia at its initial phase. Treatment aims to kill leukemia cells in the blood and bone marrow.
Learn More
Pre-B
Chronic Myeloid Leukemia (CML). Disease progresses more rapidly with higher blast cell counts in this accelerated phase.
Learn More
Pro-B
Chronic Lymphocytic Leukemia (CLL). Advanced stage with multiple affected lymphoid tissues and concerning blood counts.
Learn More
Upload Another Sample

# CHAPTER 8. CONCLUSION AND FUTURE SCOPE

## 8.1 CONCLUSION

The Blood Cancer Detection System provides an efficient and accurate method for early detection of blood cancer using deep learning techniques. The system has been rigorously tested and has demonstrated high accuracy in classification, significantly aiding medical professionals in making timely and well-informed decisions. By automating the detection process, the system minimizes human error and increases diagnostic efficiency, ultimately improving patient care outcomes.

The system's ability to process large datasets ensures scalability and adaptability for realworld medical applications. The integration of advanced machine learning techniques, such as MobileNet, enhances the detection accuracy, making the system a reliable tool for early cancer diagnosis. Additionally, the automated report generation and user-friendly interface facilitate ease of use by healthcare professionals, reducing the manual workload of medical practitioners.

With a focus on security and compliance, the system ensures that patient data remains confidential and protected against unauthorized access. The successful implementation and validation of the system indicate its potential for widespread adoption in hospitals, research laboratories, and telemedicine applications. The advancements in AI-driven healthcare solutions demonstrated by this system pave the way for future innovations in automated medical diagnostics.

## 8.2 FUTURE SCOPE:

- **Expanding Dataset**: Incorporating more diverse and larger annotated datasets for significantly improved model accuracy.
- **Real-Time Analysis**: Implementing robust real-time detection capabilities using scalable and secure cloud-based services.

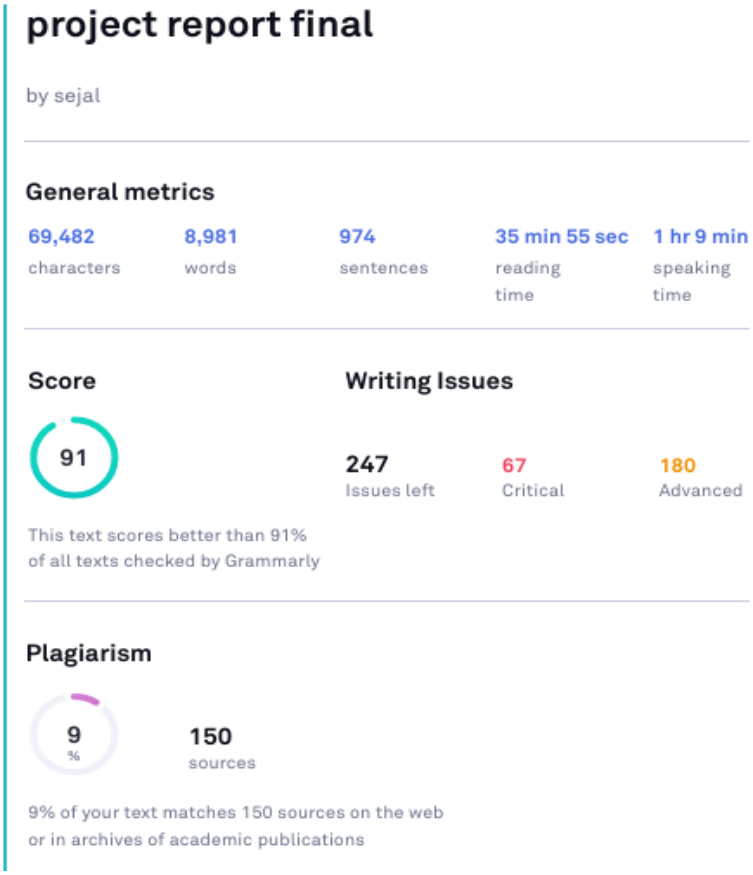
- **Multi-Class Classification**: Extending the model to accurately and efficiently detect multiple types of blood-related disorders.
- **Integration with EHR Systems**: Enabling seamless and secure integration with electronic health record (EHR) systems for efficient patient management.
- **Mobile Application**: Developing a user-friendly mobile-based application to make the system easily accessible for remote diagnostics.
- **Explainable AI (XAI) Integration**: Enhancing transparency in model decisions by providing interpretable outputs for medical professionals.
- **Personalized Diagnosis**: Leveraging patient-specific historical data to improve predictive accuracy for different demographics and medical histories.

### 8.3 Applications

The Blood Cancer Detection System has various applications, including:

- **Hospitals & Clinics**: Assisting doctors and pathologists in rapid and highly accurate blood cancer diagnosis through AI-powered analysis.
- **Research Institutions**: Supporting advanced medical research and improving AIdriven healthcare solutions and models.
- **Medical Laboratories**: Enhancing efficient automated screening and significantly reducing the workload of lab technicians.
- **Telemedicine**: Providing reliable and efficient remote diagnosis capabilities for patients in underserved and rural regions.
- **Pharmaceutical Companies**: Helping in innovative drug discovery by analyzing the progression patterns of blood cancer.
- **Public Health Initiatives**: Supporting early detection campaigns to reduce mortality rates associated with late-stage cancer diagnoses.

## 8.3 Plagiarism Report



project report final

by sejal

General metrics

| 69,482 | 8,981 | 974 | 35 min 55 sec | 1 hr 9 min |
|---|---|---|---|---|
| characters | words | sentences | reading time | speaking time |

Score

91

This text scores better than 91% of all texts checked by Grammarly

Writing Issues

| 247 | 67 | 180 |
|---|---|---|
| Issues left | Critical | Advanced |

Plagiarism

9 %

150 sources

9% of your text matches 150 sources on the web or in archives of academic publications

## REFERENCES


1. Ferreira, Fernando Rodrigues Trindade, and Loena Marins do Couto. "Using deep learning on microscopic images for white blood cell detection and segmentation to assist in leukemia diagnosis." The Journal of Supercomputing 81.2 (2025): 1-42.
2. Shehta, A.I., Nasr, M. & El Ghazali, A.E.D.M. Blood cancer prediction model based on deep learning technique. Sci Rep 15, 1889 (2025). https://doi.org/10.1038/s41598-02484475-0
3. Al-Obeidat F, Hafez W, Rashid A, Jallo MK, Gador M, Cherrez-Ojeda I and SimancasRacines D (2025) Artificial intelligence for the detection of acute myeloid leukemia from microscopic blood images; a systematic review and meta-analysis. *Front. Big Data* 7:1402926. doi: 10.3389/fdata.2024.1402926
4. Ahad, Md Taimur, et al. "Dvs: Blood cancer detection using novel cnn-based ensemble approach." arXiv preprint arXiv:2410.05272 (2024).
5. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems.
6. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). *MobileNetV2: Inverted residuals and linear bottlenecks*. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
7. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
8. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. IEEE Conference on Computer Vision and Pattern Recognition.
9. Litjens, G., et al. (2017). *A survey on deep learning in medical image analysis*. Medical Image Analysis, 42, 60-88.
10. Abbas, N., Jaffar, M. A., & Khan, S. D. (2021). *A review of leukemia detection using machine learning and deep learning*. Computers in Biology and Medicine, 132, 104296.
11. N. Saranyan, N. Kanthimathi, P. Ramya, N. Kowsalya and S. Mohanapriya, "Blood Cancer Detection using Machine Learning," *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2021, pp. 1-11A. Rehman, N. Abbas, T. Saba, S. I. U. Rahman, Z. Mehmood and H. Kolivand, "Classification

of acute lymphoblastic leukemia using deep learning", *Microsc. Res. Techn.*, vol. 81, no. 11, pp. 1310-1317, Nov. 2018.

12. S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks", *Technol. Cancer Res. Treatment*, vol. 17, Jan. 2018.
13. Y. Liu and F. Long, "Acute lymphoblastic leukemia cells image analysis with deep bagging ensemble learning" in CNMC Challenge: Classification in Cancer Cell Imaging, Singapore:Springer, pp. 113-121, 2019.
14. Hsu, Ching-Hsien, Xing Chen, Weiwei Lin, Chuntao Jiang, Youhong Zhang, Zhifeng Hao, and Yeh-Ching Chung. "Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning." Measurement 175 (2021): 109145.
15. Surya Sashank, Gundepudi V., Charu Jain, and N. Venkateswaran. "Detection of Acute Lymphoblastic Leukemia by Utilizing Deep Learning Methods." In Machine Vision and Augmented Intelligence—Theory and Applications, pp. 453-467. Springer, Singapore, 2021.
16. Karami, Keyvan, Mahboubeh Akbari, Mohammad-Taher Moradi, Bijan Soleymani, and Hossein Fallahi. "Survival prognostic factors in patients with acute myeloid leukemia using machine learning techniques." PloS one 16, no. 7 (2021): e0254976.
17. Tarek, Shahd, Hala M. Ebied, Aboul Ella Hassanien, and Mohamed F. Tolba. "White blood cells segmentation and classification using swarm optimization algorithms and multilayer perceptron." International Journal of Sociotechnology and Knowledge Development (IJSKD) 13, no. 2 (2021): 16-30.
18. Mirmohammadi, Pouria, Marjan Ameri, and Ahmad Shalbaf. "Recognition of acute lymphoblastic leukemia and lymphocytes cell subtypes in microscopic images using random forest classifier." Physical and Engineering Sciences in Medicine 44, no. 2 (2021): 433-441.
19. Killock, D. CancerSEEK and destroy — a blood test for early cancer detection. *Nat Rev Clin Oncol* **15**, 133 (2018).
20. Milan Tripathi, "Analysis of Convolutional Neural Network based Image Classification Techniques", *in Journal of Innovative Image Processing*, 2021.